

Investigating Lexical Diversity for Stylometric Analysis: A Corpus-Based Comparison of Three English Translations of The Arabian Nights' Entertainments

Nisrin Faraj Almabrok Omar*

Translation department, Faculty of Arts and Languages, University of Tripoli

Email: Nisreen.faraj.almabrok@gmail.com

<https://orcid.org/0009-0004-8700-1766>

التنوع المعجمي كأداة قياس أسلوبية: مقارنة قائمة على مدونات لغوية لثلاث
ترجمات باللغة الإنجليزية لقصص ألف ليلة وليلة
نسرين فرج المبروك عمر*
قسم الترجمة، كلية الآداب واللغات، جامعة طرابلس ، ليبيا

الملخص:

تبحث هذه الدراسة التباين في التنوع المعجمي عبر ثلاث ترجمات إنجليزية مختارة من قصص ألف ليلة وليلة. ويتمثل الهدف الرئيس في تقييم مدى قدرة مؤشرات التنوع المعجمي على الكشف عن الاختلافات في السمات المعجمية بين الترجمات الثلاث. وللحد من تأثير طول النص، قورنت النصوص الكاملة المختارة بعينات مضبوطة الطول. وتشير النتائج إلى وجود أنماط من التباين المعجمي بين الترجمات؛ إذ يرتبط ريتشارد برتون بتنوع معجمي أكبر ونسبة أعلى من المفردات الأقل شيوعًا والأكثر تعقيدًا، بينما يتسم إدوارد لين بأسلوب معجمي أكثر رسمية وتعقيدًا من الناحية البنيوية، في حين يقدم جون باين أسلوبًا معجميًا أكثر توازنًا نسبيًا. كما تكشف المؤشرات عن

درجات متفاوتة من التمايز والتقارب، ولا سيما بين المؤشرات الأقل تأثراً بتغير طول النص. وتخلص الدراسة إلى أن مقاييس التنوع المعجمي يمكن أن تسهم في الكشف عن أنماط التباين الأسلوبي المعجمي، لكنها لا تستطيع بمفردها إثبات أسلوب المترجم بوصفه بناءً متعدد الأبعاد وأكثر تعقيداً. وتخلص الدراسة أن مقاييس التنوع المعجمي، رغم قدرتها على الكشف عن أنماط من التباين الأسلوبي المعجمي، لا تكفي بمفردها لتقديم صورة شاملة عن أسلوب الترجمة. وبناءً على ذلك، توصي الدراسة أن تجمع بين مقاييس معجمية متعددة وتوظف أساليب إحصائية ملائمة، إلى جانب تطبيق منهجية النوافذ المنزلقة (sliding-window methodology) لفحص اتساق الأنماط المرصودة عبر أجزاء مختلفة من النصوص وتعزيز قابلية تعميم النتائج.

الكلمات المفتاحية: التنوع المعجمي؛ مترجم بالإنجليزي؛ ألف ليلة وليلة.

Abstract

This study investigates variation in lexical diversity across three selected English translations of *The Arabian Nights' Entertainments*. Its primary objective is to assess the extent to which lexical diversity measures can reveal differences in the lexical features of the three translations. To reduce the effect of text length, the selected full texts were compared with length-controlled samples. The findings indicate distinct patterns of lexical variation across the three translations. Richard Burton is associated with greater lexical diversity and a higher proportion of less common and more complex vocabulary, whereas Edward Lane exhibits a more formal and structurally complex lexical style, while John Payne demonstrates a relatively more balanced lexical profile. The measures also reveal varying degrees of differentiation and convergence among the translations, particularly among measures that are less sensitive to changes in text length. The study concludes that lexical diversity measures can contribute to identifying patterns of stylistic variation at the lexical level, but cannot, on their own, establish translator style as a multidimensional and more complex construct. Accordingly, the study recommends that future research combine multiple lexical measures and employ appropriate statistical methods, alongside a sliding-window methodology, to examine the consistency of observed patterns across different sections of the texts and enhance the generalizability of the findings.

Keywords: lexical diversity; English translation; Arabian Nights' Entertainments.

1. Introduction

The Arabian Nights' Entertainments is a compilation of largely Middle Eastern and Indian tales whose origins and authors remain unknown; the wide range of cultures represented in these stories suggests that the collection is a collaborative effort rather than the work of an individual author (Britannica Editors, 2026a). The work was introduced for the first time to Europe in the years 1704–1717 by the French orientalist Antoine Galland in his book *Les Mille et une nuits* (Britannica Editors, 2026b). A widely accepted proposition is that Galland's translation of *Alf Laylah wa Laylah* tends to represent a combination of medio-translation, reframing, and rewriting (Taha, 2025). Though based on the Syrian manuscript, the translation was the result of audacious additions and deletions and was therefore called by many a "French adaptation" rather than a "French translation" (ibid., 2025), a shift that is further reflected in Galland's choice of title *The Arabian Nights' Entertainments*, rather than *A Thousand and One Nights* (The Arabian Nights in European Literature, n.d.; Johnson, 2013) .

The English translation was rendered from French in 1715 and printed by Grub Street Alşıbli (2022), however, in the early 18th century, when these tales became known in Europe, they contributed to a wave of similar materials and pseudo-translations, giving rise to a literary genre known as the "Oriental tale," which flourished in various forms throughout the 18th century and beyond. Numerous scholars claimed to have identified Arabic texts from their literary research as the origins of these stories, only to discover that what they believed to be the textual evidence was in fact back-translations from French. As a result, the novel and its imitations became a structural part of the European literary landscape and continue to exert their influence to this day (The Arabian Nights in European Literature, n.d.; Alşıbli, 2022; Johnson, 2013; Britannica Editors, 2026a; Taha, 2025). Since the text exists in multiple manuscript forms and early translations involved adaptation, comparison with the Arabic source text may provide an additional basis for interpretation; however, such comparison was beyond the

scope of the present study. Nevertheless, Makdisi and Nussbaum (2008) describe the Nights as a "vertiginously unstable text" whose impact in Europe arose from numerous translations, adaptations, and imitations rather than from any single authoritative original. In addition, Al-Musawi (2021) highlights the extensive history of translation and retranslation of *The Arabian Nights Entertainments* and emphasizes the role of translators in shaping its representations across different cultural contexts. Accordingly, this study treats the translated texts as distinct literary versions of the same narrative material, shaped by historical, editorial, and cultural factors, and adopts a corpus-based approach that focuses on measuring lexical and stylometric variation across the translations themselves. The variation among translated texts and the contested status of the original is widely acknowledged in scholarship; therefore, this study does not devote extensive discussion to this issue.

1.1 Statement of the Problem

Despite the extensive study of lexical diversity in literary works, and in this subject, *Arabian Nights' Entertainments*, there is limited quantitative research that systematically applies multiple lexical diversity (LD) indices to investigate and identify the distinctive patterns of stylistic variations across multiple English translations of *Arabian Nights' Entertainments*.

Existing studies in translation and stylistics often discuss the stylistic variation of *Arabian Nights' Entertainments* qualitatively, while quantitative LD-based comparisons across these specific English versions remain methodologically underdeveloped. Consequently, the extent to which LD measures can detect and differentiate potential stylistic variation among these translations has not been sufficiently investigated.

1.2 Aims of the study

This study aims to investigate the stylistic variation among the selected English translations of *Arabian Nights' Entertainments* through the analysis of lexical diversity using specific LD indices.

1.3 Research questions

1. What patterns of LD variation can be identified across the selected translations as a manifestation of the translator's style?
2. How do different LD indices compare in describing translators' variation?

3. To what extent the researcher can utilize the lexical diversity measurements to detect translators' style?

1.4 Significance of the Study

- Contribute to addressing the existing knowledge gap in quantitative research on *One Thousand and One Nights*.
- Contributes to translation studies by exploring the potential of LD metrics to reflect stylistic variation within English translations of the same source text.
- Highlights the interpretative boundaries of the selected lexical indices in corpus-based literary translation, clarifying their scope and limitations.

2. Literature Review

Translation has always played a pivotal role in our understanding of world literature. David Damrosch highlighted this idea in his writings; according to Forsdick (2024), Damrosch argues in several of his studies that the circulation of a literary work across multiple contexts and in different languages enhances its status and lends it a universal character. In other words, translation allows texts to "spread into a wider world beyond their linguistic and cultural homeland." Furthermore, literature is complex, whether through vocabulary, metaphors, syntax, narrative structure, temporality, or other features. It would be challenging enough to capture any or all of these features, even were they to remain constant across time (Jones & Nulty, 2019). Given the complexity and multidimensionality of literary texts, quantitative approaches can complement qualitative interpretation by providing systematic means of identifying and measuring recurring patterns in their linguistic structure (Gao & Zhou, 2024, Davis & Brown, 2024).

Among the linguistic features that can be examined quantitatively, lexical diversity (LD) has received considerable scholarly attention as a measure of variation in vocabulary use. It is the subject of a significant amount of research in literature (Bestgen, 2024). Corpus-based research has demonstrated that LD is a measurable and meaningful characteristic of literary texts (ibid., 2023). When LD indices are applied within computational stylistics, LD indices contribute to the study of lexical richness (Zheng, 2025a) and genre analysis (Akbari & Jarvis, 2023), shedding light on how

vocabulary usage differentiates textual domains (Zheng, 2025b). Nevertheless, contemporary stylistics is divided into two main directions or schools, according to the classification of Pierre Guiraud: traditional stylistics, whose foundations were established by Bally, and modern stylistics, which Jakobson derived from the structuralist orientation of the Prague School. Both schools agree in defining style as the distinctive form of the text. Accordingly, research investigating differences among translated texts tends to prioritize the identification of the translator's stylistic imprint (Saad, 1993, p. 18).

Recent empirical research has put one of the stylistic assumptions to the test regarding translation universals. A study conducted by Wang et al. (2025) examined lexical and syntactic simplification in translated English using entropy analysis, disputing the common suggestion of simplification as a translation universal in translation studies (Baker, 1993, 1996). The results indicated that translated texts show greater lexical complexity, with higher wordform entropy than native texts. This entropy-based approach enhances the understanding of translational language by moving beyond traditional analyses. The study suggests that while translated texts exhibit higher lexical complexity, their syntactic complexity is similar to native texts, requiring further research across genres. Additionally, varying text types may demand different strategies for managing complexity, and entropy measurements could provide useful benchmarks for large-scale translation projects. In a similar manner, Fan et al. (2025) investigated the universal simplification hypothesis by comparing the linguistic complexity of translated texts (TTs) and non-translated texts (NTTs). Using entropy and mean dependency distance to measure lexical and syntactic complexity, respectively, the study found that TTs exhibited greater lexical complexity and at least comparable syntactic complexity to NTTs. These findings challenge the traditional assumption of universal simplification in translation studies, suggesting that translated language does not necessarily conform to a uniform pattern of simplification. The study therefore highlights the variability and complexity of translational language and provides further evidence against treating simplification as a universal characteristic of translation.

Consequently, Translation does not necessarily exhibit a uniform linguistic pattern. If translated language does not consistently conform to a universal pattern of simplification, investigating variation among individual translations becomes particularly relevant to translator-style research.

In this regard, another example is the study of "Lexical Diversity in a Literary Genre: A Corpus Study of the R̥gveda," which applied quantitative corpus analysis to examine lexical variation within a classical literary corpus. The study investigated LD within a literary corpus by examining three measures: type-token ratio (TTR), frequent content words ratio (FCWR), and hapax ratio (HR). The findings revealed that LD varied across sections of the corpus, suggesting the presence of internal stylistic variation within the same literary genre. The study therefore demonstrates the potential of LD measures for identifying patterns of linguistic variation within literary corpora (Sotov, 2009).

Beyond these applications, it is equally crucial to address the methodological limitations of LD measures. The most challenging one is text-length dependency. No single LD metric has been found to be entirely independent of text length (Zheng, 2025b). This long-standing challenge has prompted extensive efforts for nearly a century to evaluate how different indices behave under varying conditions of text length (Bestgen, 2024). The most common metric is Type-Token Ratio (TTR), where the main limitation of TTR is its dependence on text length: as a text becomes longer, the likelihood of word repetition increases, causing TTR to decrease systematically (Rosillo-Rodes et al., 2025).

Moreover, texts shorter than 500 words show considerable variability, mostly due to sample size effects rather than actual differences in lexical content. Empirical evidence shows that stability in metric values emerges between 1,000 and 2,000 words (Zheng, 2025b). Empirical studies suggest that the statistical instability inherent in LD indices is substantially reduced under conditions of length homogeneity, genre uniformity, and standardized tokenization and lemmatization procedures, thereby enhancing the internal reliability of the measure within controlled comparative designs. To overcome these length-related issues. There is evidence that MTLT generates scores that are strong indicators of the LD of language samples. Several recent

studies have strongly recommended the use of the moving average type-token ratio (MATTR) to estimate the LD of a text because it is the only length-insensitive index that can compare texts of different sizes (Bestgen, 2024; Zenker & Kyle, 2020; Kyle et al., 2023, Fergadiotis et al., 2015).

3. Methodology

3.1 Study Design

This study employs a quantitative design focusing on LD as one dimension of lexical and stylistic variation. *Arabian Nights' Entertainments* was selected as the primary text for analysis.

3.2 Data Collection

The corpus comprises three English translations of *The Arabian Nights' Entertainments* produced during the nineteenth century: Edward William Lane's translation (1838–1859), John Payne's translation (1882–1884), and Richard Francis Burton's translation (1885–1888). These translations were selected because they represent three prominent and historically distinct English versions of The Arabian Nights, produced across different periods of nineteenth-century translation history and associated with different translational approaches. Their selection therefore provides a suitable basis for examining whether LD indices can identify systematic differences in the lexical profiles of different translations of the same literary work.

The story "*The Fisherman and the Jinni*" was selected because it appears in all three translations. The complete versions of the story range from 2,089 to 2,671 tokens. In addition to the full-text analysis, a fixed-length subsample of 2,000 tokens was used to reduce the effect of differences in text length.

3.3 Data description

The data description identifies a brief overview of each translator's stated translation approach to provide context for understanding the quantitative findings.

- Richard Francis Burton (1885-1888): Literal translation imitating Arabic rhythmic prose; Chaucerian/Elizabethan English; obscure or archaic words in his attempt to maintain the flavour of the original; extensive scholarly notes. His translation was based on the Cairo text (Sandilands, 2024);

- Edward William Lane (1838–1859): Preserved authenticity; annotated Arabic customs; he altered in his translation several elements to the original text and tales, such as the representation of women and style of narration. His translation was based on the Bulaq manuscript (Lane & Lane-Poole, 1910; Van-Doorne, 2020).
- John Payne (1882–1884): Avoided complex academic transliteration systems and apostrophes/diacritical marks for proper names; rendered poetry in English metrical shape preserving external form and rhyme; focused on general sense and literary fitness using the best English idiomatic equivalents rather than strict literalism; reviewed and revised with the assistance of Richard F. Burton. His translation was based on a variorum text combining the Macnaghten, Bulaq, and Breslau editions (Payne, 1882).

3.4 Data Pre-processing

All texts were pre-processed under an identical deterministic protocol to maximize reproducibility. All three translations of the story *the fisherman* are isolated separately in Microsoft Word.

Except for Sentence Count, Average Sentence Length, Flesch-Kincaid Reading Ease, and Type-Token Ratio, the translations were retained in their raw form prior to analysis. For the remaining LD measures, the following pre-processing procedures were applied:

- All punctuation (e.g., !?.,-) is removed, as are other non-standard characters (e.g., \/* and non-UTF-8 quotes).
- All tabs, newlines (breaks), double spaces, et cetera are removed. Remaining spaces are used as word boundaries, and they are not counted.
- All letters are converted to lower case (because 'Speak,' 'SPEAK,' and 'speak' are treated as one type and three tokens).
- Possessive apostrophes and the corresponding possessive s were removed (e.g., afrits' → afrit) to prevent possessive forms from being counted as separate tokens.

Tokens were defined as orthographic words separated by whitespace following punctuation removal.

Moreover, all texts were analyzed using the LD Measurements tool under identical computational settings. The following options were applied within the analytical tool prior to the calculation of the LD indices:

- Stemming was disabled; therefore, no stemming based on the Porter algorithm was applied.
- The natural logarithm was used for Maas's a^2 , Dugast's U^2 , and Herdan's C .
- The maximum number of words analyzed per text was set to 4,000 tokens.
- Numerical results were reported to two decimal places.

3.5 Instrumental Tool

Reunecker LD Measurements tool, an online corpus-based analytical instrument was employed for quantitative data processing (Reunecker, 2017). There is evidence from previous research that this statistical tool has been utilized, including studies by Parmar et al. (2026) and Fujita & Shintani (2025).

4. Justification and Rationale of the Study

A direct comparison of LD between the Arabic source text and the English translations was not conducted. LD indices are sensitive to morphological structure and tokenization practices. Arabic relies on a root based morphological system with extensive derivation and inflection, whereas English shows comparatively lower morphological density. These structural differences directly influence the number of surface word forms and therefore affect type counts (Jasanoff & Cowgill, 2026; Wolff, 2024).

Measures such as type token ratio and related indices are influenced by inflectional variation and morphological productivity. Without systematic normalization procedures such as lemmatization or morphological segmentation, cross linguistic comparison may reflect typological differences rather than stylistic variation. For this reason, the quantitative analysis was limited to the three English translations in order to maintain methodological control and reduce structural bias.

Although the corpus size is relatively limited, previous research indicates that LD measures such as MTLT and MATTR provide stable estimates within samples exceeding 1000 tokens (Zenker & Kyle, 2020; Fergadiotis et al.,

2015). Therefore, the present corpus falls within the lower acceptable range for exploratory stylometric analysis, and length-robust measures (MTLD, MATTR, MSTTR) were included to mitigate length sensitivity associated with proportional measures such as TTR. Finally, the selected story was chosen because it exists in all three translations with comparable narrative structure, enabling controlled intralingual comparison of the translations.

5. Results and Discussion

This section presents descriptive statistical patterns of LD across the three English translations of the story "*the Fisherman and Jinni*."

5.1 Corpus Structure and Length-Standardized Comparison

To examine the stability of LD indices under varying text lengths, two analytical conditions were implemented. The first condition analyzed a length-standardized sample consisting of the first 2,000 tokens from each translation. The second condition analyzed the full text of each translation, with lengths ranging from 2,089 to 2,664 tokens. This dual design enables systematic evaluation of the sensitivity of each LD metric to text length variation. Table 1 presents the structural characteristics of the full texts.

Table 1: Corpus Size and Structural Characteristics (Full Texts)

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Tokens	2671	2089	2414
Types	799	552	654
Sentences	96	71	108
Flesch-Kincaid Reading Ease	68.6	63.38	75.33
Average sentence length	27.94 (<i>sd</i> = 17.5)	29.35 (<i>sd</i> = 23.79)	22.52 (<i>sd</i> = 15.59)
Average word length	1.3 (<i>sd</i> = 0.58)	1.34 (<i>sd</i> = 0.67)	1.29 (<i>sd</i> = 0.57)

As shown in Table 1, Burton's translation had the highest number of tokens (2,671) and types (799), followed by Payne (2,414 tokens, 654 types) and Lane (2,089 tokens, 552 types). Sentence count also varies considerably, with Payne having the highest number of sentences (108) and Lane the lowest (71). These figures show a generally consistent pattern, as longer texts provide greater opportunities for more tokens and types. Average sentence length is

highest in Lane (29.35 words), followed by Burton (27.94) and Payne (22.52). The Flesch-Kincaid Reading Ease scores indicate that Payne's translation is the most readable (75.33), followed by Burton (68.6) and Lane (63.38), suggesting a relatively higher level of reading difficulty in Lane's translation. Overall, the three translations suggest distinct patterns in the way the translators linguistically realize the same narrative. Payne tends to present the narrative in a more segmented manner, using shorter and more frequent sentences, which makes his version relatively easier to read. Lane, in contrast, constructs the narrative through fewer but longer sentences, resulting in a relatively more difficult reading experience. Burton occupies an intermediate position, with a sentence structure that falls between those of Payne and Lane and a corresponding level of readability. Thus, the three translations differ not only in how the narrative is organized at the sentence level but also in their overall ease of reading. However, further descriptive and statistical analyses are needed to examine these differences more closely, as the variation in sentence length may result from differences in syntactic complexity. Such analyses would help determine whether the observed differences reflect systematic variation in the translators' syntactic styles. Moreover, because the texts were not lemmatized and because TTR is sensitive to text length, these descriptive results alone do not establish that any translator has greater LD.

5.2 Lexical Diversity: Proportional Measures

Proportional measures quantify the relationship between unique lexical items (types) and total lexical items (tokens). These indices reflect the surface density of unique vocabulary within a text. Table 2 presents the proportional LD measures for both the length-standardized sample and the full texts.

Table 2: Proportional Lexical Diversity Measures

Investigating Lexical Diversity for Stylometric Analysis: A Corpus-Based Comparison of Three English Translations of The Arabian Nights' Entertainments

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Type-token ratio (TTR)	0.304 (full text). 0.34 (length-standardized sample).	0.264 (full text). 0.27 (length-standardized sample).	0.271 (full text). 0.3 (length-standardized sample).
Guiraud's index	15.46 (full text). 15.25 (length-standardized sample).	12.08 (full text). 11.99 (length-standardized sample).	13.31 (full text). 13.19 (length-standardized sample).
Herdan's C	0.85 (full text). 0.86 (length-standardized sample).	0.83 (full text). 0.83 (length-standardized sample).	0.83 (full text). 0.84 (length-standardized sample).
Hapax Legomena	487(60.95%) (full text). 440 (64.52%) (length-standardized sample).	313 (56.7%) (full text). 306 (57.09%) (length-standardized sample).	386 (59.02%) (full text). 365 (61.86%) (length-standardized sample).
Dis Legomena	132 (16.52%) (full text). 106 (15.54%) (length-standardized sample).	91 (16.49%) (full text). 88 (16.42%) (length-standardized sample).	107 (16.36%) (full text). 86 (14.58%) (length-standardized sample).

As shown in Table 2, Burton's translation records the highest values for TTR (0.304), Guiraud's Index (15.46), and Herdan's C (0.85) in the full texts, followed by Payne (0.271, 13.31, and 0.83, respectively) and Lane (0.264, 12.08, and 0.83). The same general pattern is observed in the length-

standardized samples, with Burton again recording the highest values (0.34, 15.25, and 0.86), followed by Payne (0.30, 13.19, and 0.84) and Lane (0.27, 11.99, and 0.83). Hapax Legomena show a similar pattern: Burton has the highest proportion in both conditions, followed by Payne and Lane. Dis Legomena, however, shows a less consistent ranking. In the full texts, the three translators have very similar proportions, whereas in the length-standardized samples Lane records the highest proportion, followed by Burton and Payne. Overall, Burton consistently occupies the highest position across the main lexical-diversity measures, while Lane generally occupies the lowest position and Payne falls between the two.

The differences between the measures can be explained by the different aspects of lexical distribution that they capture and by their different sensitivity to text length. TTR expresses the proportion of different word types relative to the total number of tokens and is therefore particularly sensitive to corpus size. The higher TTR observed for Burton indicates a greater proportion of different word types relative to the size of his text, and this ranking remains unchanged after length standardization. Guiraud's Index also places Burton above Payne and Lane in both conditions, but the relatively small change between the full texts and standardized samples indicates greater stability across corpus sizes. This is consistent with its use of the square root of the total number of tokens to reduce the effect of text length. Herdan's C produces the same general pattern, with Burton having the highest value, but it provides less differentiation between Payne and Lane, whose values are identical in the full texts and standardized samples. This difference reflects the distinct mathematical treatment of vocabulary size and text length in Herdan's measure, which uses the logarithms of types and tokens. Thus, Guiraud's Index preserves clearer differences between Payne and Lane, whereas Herdan's C produces a more compressed pattern.

Hapax Legomena provide additional information about the degree of lexical repetition. Burton has the highest proportion of words occurring only once in both corpus conditions, followed by Payne and Lane. This indicates that a larger proportion of Burton's lexical items are used only once within the text. Dis Legomena, in contrast, measures words occurring exactly twice and therefore reflects a different level of lexical repetition. The proportions are

very similar across the three translators in the full texts, while the length-standardized samples show Lane with the highest proportion, followed by Burton and Payne. This difference does not necessarily indicate greater or lower lexical diversity, as Dis Legomena focuses specifically on words occurring twice. Rather, the results show that the three translations differ in the way lexical repetition is distributed across different levels of frequency.

5.3 Lexical Diversity: Vocabulary Growth Measures

Vocabulary growth measures model the rate at which new lexical items are introduced as the text progresses. Unlike proportional measures, these indices capture the dynamics of vocabulary expansion rather than static density. Table 3 presents the vocabulary growth measures.

Table 3: Vocabulary Growth Measures

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Yule's I	7.69 (full text)	4.45 (full text)	5.38 (full text)
	10.0 (length-standardized sample)	4.6 (length-standardized sample)	6.29 (length-standardized sample)
Maas's a ²	0.02 (full text)	0.02 (full text)	0.02 (full text)
	0.02 (length-standardized sample)	0.02 (length-standardized sample)	0.02 (length-standardized sample)
Dugast's U ²	52.2 (full text)	43.91 (full text)	46.46 (full text)
	53.7 (length-standardized sample)	43.88 (length-standardized sample)	47.33 (length-standardized sample)

As shown in Table 3, Yule's I records the highest value for Burton (7.69), followed by Payne (5.38) and Lane (4.45) in the full texts. The same ranking is observed in the length-standardized samples, with Burton recording 10.00, followed by Payne at 6.29 and Lane at 4.60. Maas's a² shows virtually identical values across the three translations in both corpus conditions.

Dugast's U^2 , however, follows the same ranking as Yule's I , with Burton recording the highest value (52.20), followed by Payne (46.46) and Lane (43.91) in the full texts. This pattern is also maintained in the length-standardized samples, where Burton remains highest (53.70), followed by Payne (47.33) and Lane (43.88). Overall, Burton occupies the highest position in Yule's I and Dugast's U^2 , Payne generally occupies an intermediate position, and Lane records the lowest values, while Maas's a^2 provides virtually no differentiation among the three translations.

The results can be explained by the different properties of lexical frequency captured by each measure. Yule's I is based on the distribution of word frequencies within the text and is particularly sensitive to the occurrence of frequently repeated lexical items. Its higher values for Burton, followed by Payne and Lane, therefore indicate differences in the pattern of lexical repetition across the three translations. Importantly, this ranking remains unchanged in the length-standardized samples, suggesting that the observed differences in frequency distribution are not solely attributable to differences in text length. Dugast's U^2 similarly takes lexical repetition into account, but approaches lexical diversity through the relationship between vocabulary size and the frequency distribution of word types. Its consistent ranking of Burton, Payne, and Lane, respectively, in both corpus conditions therefore provides additional evidence of differences in the distribution of lexical items across the translations.

5.4 Lexical Diversity: Sequential Length-Robust Measures

Sequential measures were developed to address the limitations of traditional indices, particularly their sensitivity to text length and uneven lexical distribution. These indices calculate lexical diversity across sequential segments of the text, providing more stable estimates that are largely independent of total text length. Table 4 presents the sequential length-robust measures.

Table 4: Sequential Length-Robust Measures

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Mean segmental TTR (MSTTR)	0.87 (full text)	0.84 (full text)	0.87 (full text)
	0.88 (length-standardized sample)	0.85 (length-standardized sample)	0.86 (length-standardized sample)
Moving average TTR (MATTR)	0.87 (full text)	0.84 (full text)	0.86 (full text)
	0.87 (length-standardized sample)	0.84 (length-standardized sample)	0.86 (length-standardized sample)
Measure of textual lexical diversity (MTLD)	76.91 (full text)	50.88 (full text)	66.35 (full text)
	78.16 (length-standardized sample)	54.39 (length-standardized sample)	65.21 (length-standardized sample)

As shown in Table 4, the three lexical-diversity measures generally produce a consistent ranking across the two corpus conditions. For MSTTR, Burton and Payne record the highest value in the full texts (0.87), followed by Lane (0.84), while the length-standardized samples place Burton first (0.88), followed by Payne (0.86) and Lane (0.85). MATTR shows the same ranking in both conditions, with Burton recording 0.87, followed by Payne at 0.86 and Lane at 0.84 in both the full texts and the length-standardized samples. MTLD likewise produces a consistent ranking, with Burton recording the highest value in both conditions (76.91 in the full text and 78.16 in the length-standardized sample), followed by Payne (66.35 and 65.21, respectively) and Lane (50.88 and 54.39, respectively). Overall, Burton occupies the highest position across the three measures, Payne generally occupies an intermediate position, and Lane records the lowest values.

The consistency of these results can be attributed to the design of the three measures, which reduce the influence of overall text length in different ways.

MSTTR calculates TTR across segments of equal length and then averages the resulting values, thereby limiting the effect of differences in corpus size. MATTR similarly calculates TTR across successive moving windows and averages the results, making the measure less dependent on the total length of the text. MTLD assesses the extent to which lexical diversity is maintained across the text rather than relying on a single TTR value for the entire corpus. Consequently, the persistence of the same general ranking after length standardization is important because it indicates that the higher lexical-diversity values observed for Burton are not simply a consequence of his longer text. The particularly stable results for MATTR and MTLD provide further support for the consistency of the observed differences, although the small changes in the absolute values show that corpus condition can still have a limited effect on individual measurements.

6.5 Lexical Complexity and Frequency Characteristics

Lexical complexity measures provide insight into the morphological and frequency-based characteristics of the vocabulary employed by each translator. Table 5 presents the lexical complexity and frequency indicators.

Table 5: Lexical Complexity and Frequency Characteristics

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Mean word length (MWL)	4.03 (<i>sd</i> = 1.93) (full text)	4.06 (<i>sd</i> = 2.09) (full text)	3.93 (<i>sd</i> = 1.86) (full text)
	4.01 (<i>sd</i> = 1.9) (length-standardized sample)	4.05 (<i>sd</i> = 2.09) (length-standardized sample)	3.9 (<i>sd</i> = 1.85) (length-standardized sample)
Mean word frequency (MWF)	3.29 (full text)	3.78 (full text)	3.69 (full text)
	2.93 (length-standardized sample)	3.73 (length-standardized sample)	3.39 (length-standardized sample)
Mean Zipf frequency	3.633 (average) (full text)	3.731 (average) (full text)	3.784 (average) (full text)
	3.649 (average) (length-standardized sample)	3.74 (average) (length-standardized sample)	3.811 (average) (length-standardized sample)
Mean log difficulty	12.358 (full text)	12.132 (full text)	12.01 (full text)
	12.322 (length-standardized sample)	12.111 (standardized sample)	11.949 (length-standardized sample)

Mean word length (MWL) showed Lane having the longest words (4.05–4.06 characters), followed by Burton (4.01–4.03) and Payne (3.90–3.93). This pattern remains stable across both corpus conditions. The higher MWL indicates that Lane's translation contains longer lexical items on average, although word length alone does not necessarily imply greater morphological or lexical complexity. It may also reflect differences in morphological structure, lexical choice, register, compounding, and the distribution of short function words.

Mean word frequency (MWF) shows a different pattern. Burton records the lowest values (2.93 in the length-standardized sample; 3.29 in the full text), followed by Payne (3.39; 3.69), while Lane records the highest values (3.73; 3.78). Thus, according to the frequency measure used, Burton's vocabulary has a lower average frequency, whereas Lane's vocabulary has a higher average frequency. The combination of higher MWL and higher MWF for Lane indicates that word length and lexical frequency capture different aspects of the translators' lexical profiles. For example, Lane may employ longer but relatively frequent words, including established polysyllabic or formal terms, whereas Burton may use shorter but less frequent or archaic lexical items. Thus, longer words do not necessarily correspond to lower lexical frequency.

Zipf frequency gives an approximate score for the prevalence of a word in a language, where a higher Zipf score means the word is more common and a lower Zipf score means the word is less common. However, in the results, Mean Zipf frequency shows Burton with the lowest values (3.649 in the length-standardized sample; 3.633 in the full text), followed by Lane (3.740; 3.731) and Payne (3.811; 3.784). Since lower Zipf scores generally indicate lower word frequency, these results suggest that Burton uses less frequent lexical items on average than Lane and Payne.

Mean log difficulty follows a different but complementary pattern, with Burton recording the highest values (12.322 standardized; 12.358 full text), followed by Lane (12.111; 12.132) and Payne (11.949; 12.01). Higher log difficulty values reflect lower probabilistic expectancy and greater lexical rarity within standard English reference corpora. This indicates that Burton's vocabulary is characterized by higher concentration of infrequent, archaic, or non-standard lexical items.

5.6 Compression-Based and Distributional Measures

Compression-based measures evaluate lexical redundancy and informational density using information-theoretic models. These indices reflect the predictability and compressibility of lexical sequences. Table 6 presents the compression-based measures.

Table 6: Compression-Based Measures

Statistic Type	Richard F. Burton	Edward William Lane	John Payne
Compression rate (GZ)	0.61 (full text)	0.64 (full text)	0.63 (full text)
	0.6 (length-standardized sample)	0.63 (length-standardized sample)	0.62 (length-standardized sample)
Compression rate (LZW)	0.68 (full text)	0.67 (full text)	0.68 (full text)
	0.66 (length-standardized sample)	0.67 (length-standardized sample)	0.66 (length-standardized sample)
Hirsch-Popescu Point (HPP)	19 (full text)	19 (full text)	21 (full text)
	15 (length-standardized sample)	18 (length-standardized sample)	18 (length-standardized sample)

The compression-based measures show some variation across the three translations, although the differences are relatively small. For GZ compression, Lane records the highest values (0.64 full text; 0.63 length-standardized sample), followed by Payne (0.63; 0.62), while Burton records the lowest values (0.61; 0.60). LZW compression shows a different pattern: Payne and Burton record the same values (0.68; 0.66), while Lane shows slightly lower values (0.67 in both corpus conditions).

HPP also shows some differentiation, with Payne recording the highest value in the full text (21), while Lane and Burton both record 19. In the length-standardized samples, Payne and Lane record 18, whereas Burton records the lowest value (15). Overall, these results indicate that the three translations do not follow a single consistent ranking across the compression-based measures. For example, Lane has the highest GZ value but a lower LZW value than Payne and Burton, while Burton has the lowest GZ value but the same LZW value as Payne. This suggests that the lexical or textual differences captured by these measures depend on the specific property assessed by each index.

6. Conclusion

This study investigated stylistic variation across three English translations of *Arabian Nights' Entertainments* by examining lexical diversity through a range of LD indices. The findings revealed consistent differences among the translations, with Burton generally showing the highest LD profile, Payne occupying an intermediate position, and Lane generally showing lower LD values. These patterns remained largely stable in the length-standardized samples, suggesting that the observed differences cannot be attributed to text length alone.

The comparison of the measures also showed that different LD indices capture different aspects of lexical variation and do not necessarily produce identical degrees of differentiation. Measures such as MATTR, MSTTR, and MTLTD provided relatively stable patterns across corpus conditions, while measures based on specific frequency levels, such as Hapax and Dis Legomena, revealed additional differences in lexical repetition. Thus, the findings demonstrate that LD can serve as a useful quantitative dimension for detecting variation in translator style, but no single measure can provide a complete representation of style.

Overall, the study contributes to quantitative research in translation studies by demonstrating the potential of LD measures to identify systematic differences among translations of the same narrative. At the same time, the findings indicate that LD represents only one dimension of translation style. A more comprehensive investigation would benefit from combining multiple lexical, syntactic, and textual measures with appropriate statistical analysis and sliding-window methodology along with lemmatization approach to examine the consistency of observed patterns across different parts of the texts.

7. Limitations of the study

The present findings should be interpreted within the following limitations:

1. Sample size: Although the corpus falls within the acceptable range for exploratory stylometric analysis, larger samples would increase statistical robustness.
2. Single story analysis: The study is limited to one story from the Arabian Nights' collection; future research should extend the analysis to multiple

stories to assess consistency of translator profiles.

3. Lexical focus: The analysis captures only one lexical dimension of style and does not account for syntactic embedding, rhetorical configuration, discourse organization, or pragmatic modulation.

8. Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the author used AI tools in order to improve language clarity, readability, the organization of the manuscript, and to identify the logical relationships among the findings. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

9. References

- Akbary, M., & Jarvis, S. (2023). Lexical diversity as a predictor of genre in TV shows. *Digital Scholarship in the Humanities*, 38(3), 921–936.
<https://doi.org/10.1093/llc/fqad004>
- Al-Musawi, M. J. (2021). The Arabian nights in contemporary world cultures. In Cambridge University Press eBooks.
<https://doi.org/10.1017/9781108593847>
- Alşıbli, İ. (2022). ألف ليلة وليلة في الأدب الأوروبي. *Bingöl Üniversitesi İlahiyat Fakültesi Dergisi*, (19), 196–211.
<https://doi.org/10.34085/buifd.1061535>
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–252). *John Benjamins*.
- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation*:

- Studies in language engineering, in honour of Juan Sager (pp. 175-186). *John Benjamins*.
- Bestgen, Y. (2024). Measuring Lexical Diversity in Texts: The Twofold Length Problem. *Language Learning*, 74(3), 638–671.
<https://doi.org/10.1111/lang.12630>
- Britannica Editors (2026a). The Thousand and One Nights. *Encyclopedia Britannica*. <https://www.britannica.com/topic/The-Thousand-and-One-Nights>
- Britannica Editors (2026b). *Antoine Galland*. *Encyclopedia Britannica*.
<https://www.britannica.com/biography/Antoine-Galland>
- Davis, A., & Brown, J. (2024). The paradigm of literary work analysis from the perspective of corpus stylistics. *www.cambridgepublish.com*.
<https://doi.org/10.62852/cahr/2024/59>
- Fan, L., Yao, Y., Xie, R., Sio, C., & Cheung, A. K. F. (2025). Testing the simplification universal in game localization: a quantitative comparison of linguistic complexity in black myth, Sekiro, and RDR2. *Humanities and Social Sciences Communications*, 13(1).
<https://doi.org/10.1057/s41599-025-06424-0>
- Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric Evaluation of Lexical Diversity Indices: Assessing Length Effects. *Journal of speech, language, and hearing research : JSLHR*, 58(3), 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280
- Forsdick, C. (2024). عرض موجز: الإستمولوجيات الترجمية في الأدب العالمي. *Encounters in Translation*, 2. <https://doi.org/10.35562/encounters-in-translation.570>
- Fujita, T., & Shintani, N. (2025). Evaluating a task-based language teaching course for low-proficiency learners in ESP classrooms in Japan. *Language Teaching Research*.
<https://doi.org/10.1177/13621688251324376>
- Gao, Y., & Zhou, G. (2024). Revealing the translator's style: A corpus-based study of english translations of *Mencius*. *PLoS ONE*, 19(7), e0305894. <https://doi.org/10.1371/journal.pone.0305894>

- Jasanoff, J.H., Cowgill, W. (2026, January 14). *Indo-European languages*. *Encyclopedia Britannica*.
<https://www.britannica.com/topic/Indo-European-languages>
- Olcott, F. J. (Ed.). (1913). *The Arabian Nights' Entertainments: based on a translation from the Arabic by Edward William Lane*. Henry Holt and Company.
<https://archive.org/download/arabiannightsent00lane/arabiannightsent00lane.pdf>
- Johnson, R. C. (2013). Beautiful Infidels: The Western Travels of the Arabian Nights. *Eighteenth-Century Studies*, 46(3), 434–439.
<https://doi.org/10.1353/ecs.2013.0024>
- Jones, E., & Nulty, P. (2019). Quantitative measures of lexical complexity in modern prose fiction. *Digital Scholarship in the Humanities*, 34(4), 914–937. <https://doi.org/10.1093/llc/fqz020>
- Kyle, K., Sung, H., Eguchi, M., & Zenker, F. (2023). Evaluating evidence for the reliability and validity of lexical diversity indices in L2 oral task responses. *Studies in Second Language Acquisition*, 46(1), 278–299. <https://doi.org/10.1017/s0272263123000402>
- Lane, E. W. (Trans.) & Lane-Poole, S. (Rev.). (1910). *The thousand and one nights* (Vol. 16, Harvard Classics). P. F. Collier & Son.
<https://nadianote.files.wordpress.com/2011/05/1001-nights.pdf>
- Makdisi, S., & Nussbaum, F. (2008). The Arabian Nights in historical context. In *Oxford University Press eBooks*.
<https://doi.org/10.1093/acprof:oso/9780199554157.001.0001>
- Payne, J. (1882). *The book of the thousand nights and one night* (Vol. 1). Villon Society for Private Circulation.
<https://archive.org/details/arabiannights01paynuoft>
- Reunecker, A. (2017). *Lexical Diversity Measurements*. Retrieved 24 February, 2026, from <https://www.reunecker.nl/files/ld>.
- Rosillo-Rodes, P., Miguel, M. S., & Sánchez, D. (2025). Entropy and type-token ratio in gigaword corpora. *Physical Review Research*, 7(3).
<https://doi.org/10.1103/rxxz-lk3n>
- Parmar, R. P., Daulat, S. R., Shah, R., Brady, T. T., Montague, M., & Roth, C. (2026). Readability, Accuracy, and Lexical Diversity of New

- ChatGPT Models for Common Carpal Tunnel Syndrome Questions. Hand (New York, N.Y.), 15589447251415391. Advance online publication. <https://doi.org/10.1177/15589447251415391>
- Ping, Y., & Wang, B. (2023). Retranslated Chinese classical canon Journey to the West: a stylometric comparison between Julia Lovell's retranslation and Arthur Waley's translation. *Digital Scholarship in the Humanities*, 39(1), 308–320. <https://doi.org/10.1093/llc/fqad085>
- Rosillo-Rodes, P., Miguel, M. S., & Sánchez, D. (2025). Entropy and type-token ratio in gigaword corpora. *Physical Review Research*, 7(3). <https://doi.org/10.1103/rxxz-lk3n>
- Saad, M. (1993). *Fi al-nass al-adabi: Dirasa uslubiyya ihsa'iyya* (2nd ed.) [In Arabic]. *Ein for Human and Social Studies*. <https://archive.org/details/66-1411> (Original work published 1990)
- Sandilands, B. (2024). *Sir Richard Francis Burton, translator of The Book of the Thousand Nights and a Night* [PDF]. Università Ca' Foscari Venezia. <https://open.unive.it/hitrade/books/SandilandsBurton.pdf>
- Sotov, A. (2009). Lexical Diversity in a Literary Genre: A Corpus Study of the RGVEDA. *Literary and Linguistic Computing*, 24(4), 435–447. <https://doi.org/10.1093/llc/fqn044>
- Taha, N. M. I. M. (2025). Diamond in the Rough and the Oriental Myth: Shaping Arab Cultural Identity in Early Translations of Alf Laylah wa Laylah. *Humanitarian and Natural Sciences Journal*, 6(5). <https://doi.org/10.53796/hnsj67/23>
- The Arabian Nights in European Literature - An anthology - University of Kent*. (n.d.). University of Kent. <https://www.kent.ac.uk/ewto/projects/anthology/index.html>
- Van Doorne, I. (2020). *(Un)Veiling the Orient: A close analysis of Edward William Lane's Arabian Nights' Entertainments* (Bachelor's thesis, University of Amsterdam). University of Amsterdam Repository. <https://www.scribd.com/document/713027423/Un-veiling-The-Orient-A-Close-Analysis>
- Wang, Z., Cheung, A. K. F., Xu, H., & Liu, K. (2025). Assessing lexical and syntactic simplification in translated English with entropy

- analysis. *Humanities and Social Sciences Communications*, 12(1). <https://doi.org/10.1057/s41599-025-05562-9>
- Wolff, H.E. (2024, November 29). *Afro-Asiatic languages*. *Encyclopedia Britannica*. <https://www.britannica.com/topic/Afro-Asiatic-languages>
- Zenker, F., & Kyle, K. (2020). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, 47, 100505. <https://doi.org/10.1016/j.asw.2020.100505>
- Zheng, W. (2025a). Lexical richness viewed through lexical diversity, density, and sophistication. *Digital Scholarship in the Humanities*, 40(2), 692–708. <https://doi.org/10.1093/llc/fqaf023>
- Zheng, W. (2025b). Text length requirements for stable and genre-sensitive lexical diversity measurement. *Cogent Arts & Humanities*, 12(1). <https://doi.org/10.1080/23311983.2025.2584418>
- إنعمان, إنعمان, Naaman, E. (2024). The Arabian Nights in Contemporary World Cultures: Global Commodification, Translation, and the Culture Industry, written by Muhsin J. Al-Musawi. المركز: مجلة الدراسات العربية, 3(1), 176-170. <https://doi.org/10.1163/27728250-12340041>